

COMMUNIQUE DE PRESSE

Kyutai dévoile *Unmute*, une technologie de pointe pour doter les IA de l'ouïe et de la parole

Paris, 28 Mai 2025. L'an dernier, Kyutai dévoilait Moshi, la toute première IA vocale temps réel, offrant une fluidité encore inégalée dans l'interaction. Kyutai poursuit avec *Unmute* son exploration de l'IA vocale : *Unmute* transfère instantanément de l'écrit à l'oral les capacités avancées des grands modèles de langue, comme le raisonnement ou le guidage par prompting.

Unmute est une technologie modulaire qui permet de transformer instantanément n'importe quel modèle textuel en IA vocale temps réel, grâce à des modules de transcription et de synthèse de la parole développés par Kyutai. La « personnalité » et le comportement de l'agent sont pilotables par prompts textuels, tandis que sa voix est simplement guidée par quelques secondes d'échantillons d'audio. Avec *Unmute*, la laborieuse interaction textuelle avec une IA puissante fait place à une interaction vocale fluide et naturelle.

Unmute transcrit précisément et à la volée la parole de l'utilisateur, tout en déterminant si celui-ci a terminé de parler ou marque simplement une pause en cours de phrase. La latence est donc très faible, sans interruption de l'utilisateur. La réponse textuelle générée par le modèle de langue est transmise au synthétiseur vocal d'*Unmute* qui démarre sans attendre la fin de cette réponse. La latence du système complet est ainsi réduite le plus possible.

Pour Kyutai, l'interaction des humains avec et via les machines doit passer par la voix, pour être fluide, inclusive, riche de nuances et d'émotions, tout en s'appuyant sur des modèles personnalisables, légers et puissants. Avec *Unmute*, Kyutai poursuit l'exploration dans cette voie. Acteur engagé de la science ouverte, Kyutai partagera dans les prochaines semaines la technologie d'*Unmute*, comme pour tous ses projets précédents. *Unmute* peut d'ores et déjà être testé sur unmute.sh et sera présent à VivaTech.

A propos Kyutai est un laboratoire à but non lucratif dédié à la recherche ouverte en IA, fondé fin 2023 par le groupe Iliad, CMA CGM et Schmidt Sciences. Sa mission est de repousser les frontières de l'IA au bénéfice de tous. Dirigée par des scientifiques de premier plan, l'équipe compte aujourd'hui plus de 20 membres, dont des étudiants en master et en doctorat, la formation de nouveaux talents faisant intrinsèquement partie de ses missions. Kyutai œuvre à développer la multimodalité, la modularité, l'adaptabilité, la sobriété et l'interactivité des IA modernes. En moins de 18 mois, le laboratoire a conçu et partagé deux premières mondiales : Moshi, premier modèle conversationnel vocal sans latence ni tours de parole imposés, et Hibiki, premier modèle vocal de traduction simultanée avec préservation de la voix à fonctionner sur un téléphone mobile ; D'autres modèles marquants également : Mimi, codec audio aux performances vocales inégalées, Helium, modèle textuel très léger maîtrisant les 24 langues officielles de l'UE et Moshivis, modèle vocal capable de converser autour de contenus visuels.

Nous suivre [web] kyutai.org [X] [@kyutai_labs](https://twitter.com/kyutai_labs) [Bluesky] [@kyutai-labs](https://bsky.app/profile/kyutai-labs) [In] [kyutai-labs](https://www.instagram.com/kyutai_labs)
Contact [email] presse@kyutai.org

**Kyutai unveils *Unmute*, a cutting-edge technology
to make any text language model hear and speak**

Paris, 28 May 2025. Last year, Kyutai unveiled Moshi, the first real-time spoken dialogue AI model, offering unparalleled fluidity in interaction. With *Unmute*, Kyutai continues its exploration of voice AI: *Unmute* instantly transfers the extended abilities of large language models, such as reasoning and prompting guidance, from written to spoken language.

Unmute is a uniquely-modular voice AI. It empowers any text language model with voice, instantly, by wrapping it with Kyutai's new real-time speech-to-text (STT) and text-to-speech (TTS). The model's "personality" and behavior are controlled by text prompts, while its voice is simply specified by a few seconds of audio samples. With *Unmute*, contrived text-based interaction with a powerful language model is turned into fluid voice-based interaction.

Unmute's real-time STT is accurate and includes a semantic voice activity detection which predicts whether the user is actually finished speaking or is just pausing mid-sentence. This ensures very low latency, without interrupting the user. The text response generated by the language model is passed to *Unmute's* TTS, which is also real-time, starting to generate speech even before the text response is completed. This minimizes the entire system's latency.

For Kyutai, human interaction with and through machines must be voice-based, to be fluid, inclusive, rich in nuance and emotion, while relying on customizable, lightweight and powerful models. With *Unmute*, Kyutai continues its exploration in this direction. A committed player in open science, Kyutai will open-source *Unmute's* technology in the coming weeks, as done with all its previous projects. *Unmute* can already be tested on `unmute.sh` and will be presented at VivaTech.

About us Kyutai is a non-profit laboratory dedicated to open research in AI, founded in November 2023 by the Iliad Group, CMA CGM and Schmidt Sciences. Its mission is to push the frontiers of artificial intelligence for the benefit of all. Led by prominent founding scientists, the team now comprises more than 20 members, including MSc and PhD students since training new talents is also part of Kyutai's missions. Kyutai strives to endow foundation models with multimodality, modularity, adaptability, efficiency and interactivity. In fewer than 18 months the lab has developed and shared two world-premieres: Moshi, the first spoken-dialogue system with no latency and no imposed speaker turns, and Hibiki, the first simultaneous voice-to-voice translation model with voice preservation to run on a mobile phone; Other groundbreaking projects from Kyutai include Mimi, a neural audio codec with unmatched performance on speech, Helium, a lightweight language model mastering the 24 official languages of the EU, and MoshiVis, a speech model that can converse, among other things, about visual content.

Follow us [web] kyutai.org [X] [@kyutai_labs](https://twitter.com/kyutai_labs) [Bluesky] [@kyutai-labs](https://bsky.app/profile/kyutai-labs) [In] [kyutai-labs](https://www.youtube.com/channel/UCv3v3v3v3v3v3v3v3v3v3v3)
Contact us [email] press@kyutai.org